Vision-Based Semi-Supervised Homecare with Spatial Constraint

Tianqiang Liu[1], Hongxun Yao[1], Rongrong Ji[1], Yan Liu[1], Xianming Liu, Xiaoshuai Sun[1], Pengfei Xu[1], and Zhen Zhang[1]

[1] Department of Computer Science, Harbin Institute of Technology,
No. 92 West Dazhi Street, Harbin, P. R. China, 150001
{tqliu, yhx, rrji, lyan, liuxianming, xssun, pfxu, zzhang}@vilab.hit.edu.cn

Abstract. Vision-based homecare system receives increasing research interest owing to its efficiency, portability and low-cost characters. This paper presents a vision-based semi-supervised homecare system to automatically monitor the exceptional behaviors of self-helpless persons in home environment. Firstly, our proposed framework tracks the behavior of surveilled individual using dynamic conditional random field tracker fusion, based on which we extract motion descriptor by Fourier curve fitting to model behavior routines for exception detection. Secondly, we propose a Spatial Field constraint strategy to assist SVM-based exception action decision with a Bayesian inference model. Finally, a novel semi-supervised learning mechanism is also presented to overcome the exhaustive labeling behavior in previous works. Experiments over home environment video dataset with five normal and two exceptional behavior categories shows the advantage of our proposed system comparing with previous works.

Keywords: Intelligent homecare, semi-supervised learning, tracking, action detection, surveillance

1 Introduction

Automatic accident monitoring for self-helpless people at home is a promising research direction in multimedia application. Such people include elderly, children, handicapped, and other individuals who behave inconveniently. There is an emergent need for monitoring accidents of such people in the case that they live alone. These accidents include behaviors such as fall, unusual squatting, twitching, or a long period of inactivity in low frequency visited places, which are serious and often lead to injury, restricted activities, fear, or even death. Survey in [4] concludes the main reasons of bedridden for elderly people include apoplectic ictus, decrepitude, falls, and fractures [4]. Thus, the detections of such accidents in time can help the elderly to obtain in-time medical treatments. By providing intelligent, affordable, usable and expandable integration of home automation devices, homecare system can facilitate remote interaction with caregivers to provide safety and security to assist the elderly or disabled.

Y.-M.R. Huang et al. (Eds.): PCM 2008, LNCS 5353, pp. 416–425, 2008.

Homecare systems based on sensor network or wireless facility have been paid much attention by former researchers [7, 8, 9]. But considering the high expense of that system, the vision-based homecare system has been raising more research interest recently, because of its lower cost: Charif [3] summarized individual activity based on the salient region observed by cameras and introduce the action duration time in abnormal inactivity detection. Lin [2] introduced a rule-based algorithm in which also based on action duration time to detect fall incident in compressed domain. However, in our consideration, vision-based homecare systems [1, 2, 3] are far from maturity due to three following reasons:

(1) Current works on vision-based homecare system [1,2,3] mostly bases on rules which work only toward one action but cannot work to the others. For example, system oriented at fall incident detection cannot detect the twitching event [2].

(2) The particularity of homecare application environment has not been elaborately investigated. Although researcher in [3] has noticed this case (in which the analysis of the occupant's frequent-appearing region is leveraged), there still remain crucial problems unresolved: In [3], all the actions happen in the unusual active regions may be recognized as abnormal whether it is really abnormal or not. This is because different accidents may happen in different places, and the low-frequently-visited regions are diverse for different accidents.

(3) Fusion between the motion pattern analysis and spatial information has been not investigated: Current systems usually emphasize one but overlook another [1, 2, 3].

To address these problems, we propose a vision-based homecare system which merits in elaborately considerations of both scene spatial information and semi-supervised sample labeling. We present a Spatial Field constraint to restrict the spatial occurrence of behaviors to improve detection effectiveness. Also, our system only needs limited labeling to conduct exceptional behavior detection, which is achieved by semi-supervised labeling propagation to extend our training set to include highly-trusted daily shots. The systematic overview of the proposed algorithm is as follows:

Considering the surveilled individuals are usually self-helpless who behave inconveniently, we exclude some fierce actions, and define squatting down, standing up, lying down, walking and sitting down as normal actions, and falling down and twitching as exceptional actions (Shown in Figure 1).

We employ dynamic conditional random field (DCRF) based fusion strategy [5] of particle filter and mean shift trackers to track the surveilled. Then, we employ Fourier fitting to extract the motion patterns, which can acquire information both of the motion intensity and frequency according to the tracked target. Consequently, we adopt SVM classifier to analyze the motion patterns. To address the difficulties in multi-action classification, we introduce Spatial Field constraint into the action classification, and fuse motion & spatial decisions based on Bayesian inference model. Finally, the semi-supervised mechanism in SVM is introduced into the system which can retrain the classifier by automatic annotating highest confident query samples resulted from Bayesian inference. The system flow chart is as in Figure 2.

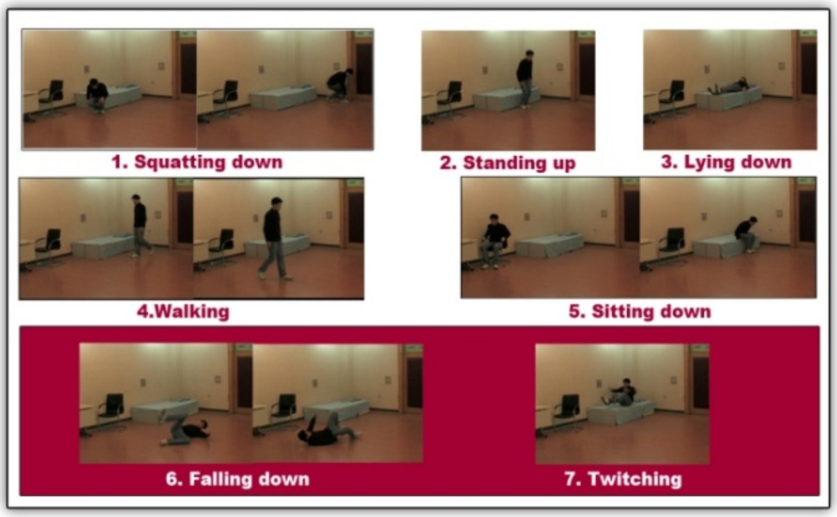


Fig. 1. Action categories definition

The rest of this paper is organized as follows: Section 2 presents our behavior tracking and motion description algorithm; Section 3 presents our action classification algorithm. Experiments are shown in Section 4. Finally this paper concludes in Section 5 and discusses our future research directions.

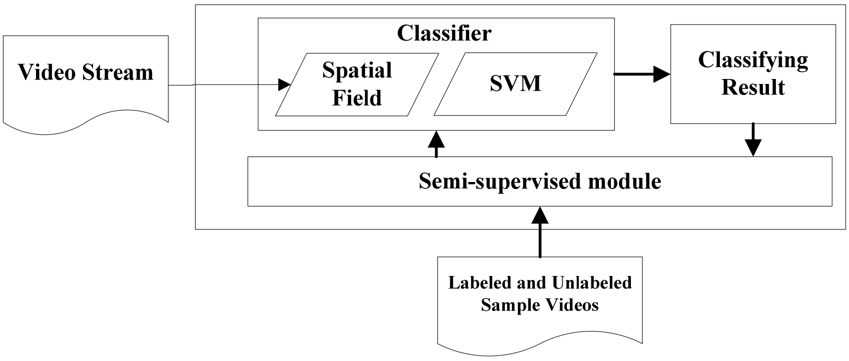


Fig. 2. Framework of Semi-Supervised homecare system

2 Motion Representation

Motion Tracking: Generally speaking, we assume the camera has been keeping static in a homecare scenario which is identical to [3], and the position and orientation to minimize occlusion of the person by furniture are selected. Differently, we do not choose the overhead orientation of view [3] because it's difficult to descript the motion patterns in this case.

The first step is tracking of the surveilled. The algorithm is required to be robust enough for the noise produced by mild change of illumination and could adjust the size of tracking window according to the person's action. We employ tracking strategy in [5] for our case, and construct a rectangular window in each frame including the surveilled person exactly: The tracker based on dynamic conditional random field (DCRF) fuses particle filter and mean shift with Bayesian inference. Then, we determine the state (foreground or background) of each foreground pixel according to all the past states and the current state. Comparing with pure particle filter, this strategy merits in combining the motion information and the hue characteristic of the target together. The experimental result is very satisfying for the failure rate of tracking is 9 out of 280.

Motion Descriptor: The training mechanism requires us to represent different actions in the uniform format. Firstly, we should select an effective feature to represent the actions. We have discovered that the the ratio of tracking window's width to height is representative in describing the motion pattern in our experiments, and thus the curve generated by frame in our system can be represented as follows:

$$R(t)=\frac{w(t)}{h(t)} \tag{1}$$

in which w(t) and h(t) represent tracking window's width and height in moment t respectively, and a t-R(t) (T-R) curve is generated for current motion pattern by calculate R values for several consecutive frames. Then we apply the Fourier series fit the motion curve and leverage the top 5 low-frequency Fourier coefficients in motion description, which acquires a 10-D feature vector as follows:

$$f(t) = a_0 + \sum_{k=1}^{\infty}(a_k \cos kt + b_k \sin kt) \tag{2}$$

3 Action Classification

The key challenge of action classification is that the T-R curves of some action categories acquired in section 2 are similar to each other, and so do the coefficients extracted by Fourier fitting. For example, sitting down and squatting down may exhibit high similarity in temporal field. The essence of our approach is to fuse the spatial information into Fourier-based SVM action classifications using Bayesian inference model.

Motion Classification: Support vector machine (SVM) [6] is used to train the action recognizer. The 10-D T-R feature is fed as feature vector into SVM. We employ the RBF kernel to map training vectors into a high dimensional feature space for classification. The SVM model returns probability form of result.

Spatial Inference: Our objective here is to adjust the classification results for some similar motion patterns easily classified incorrectly by SVM. The difficulty also exists in other classifiers to resolve the multi-action classification problem. Considering the specific environment of home shown in Figure 3, the spatial information serves as a strong feature to distinguish actions from one another. This can be demonstrated by observing specific actions' occurring locations: For example, sitting down occurs most frequently at regions at bedsides or chairs, and lying down happens usually only in bed region. This principle also makes effect to the accidents we aim to detect: falling down usually occurs on floor, nor at chair or bed regions, and twitching often near some electrical sources. Therefore, we can define Spatial Field for action *a* as follows: the specific location within the video frames which coordinate with the action *a*.

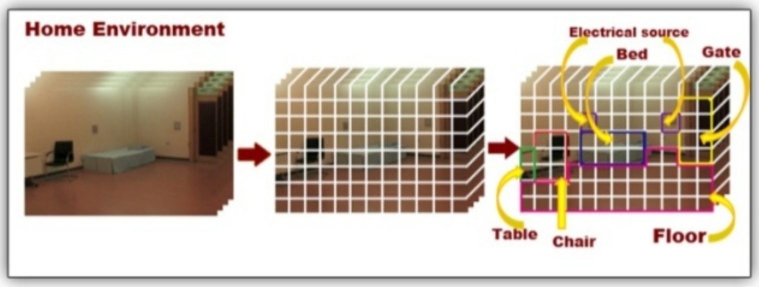


Fig. 3. Normal bedroom layout

Spatial & Motion Bayesian Fusion: The key point of our approach is to effectively integrate the Spatial Field and motion pattern analysis by SVM together. We define this advanced SVM as SFSVM in our system. The conditional independence assumption is made between the Spatial Field and motion pattern analysis, based on which we leverage Bayesian combination to infer the accident occurring probability. Firstly, the video frames are segmented into m slices. m corresponds with the camera orientation and room space which is selected as 12×8 according to experiment in our system as in Figure 4. Then, the posterior probability of each action occurring in each slice is calculated as:

$$P(e^k \mid i) = \frac{P(i \mid e^k)P(e^k)}{\sum_{j=1}^{n} P(i \mid e^j)P(e^j)} \tag{3}$$

where P(e^k), which can be acquired from probability based SVM, represents prior probability of *a* classified as k^{th} category action where $k \in \{1,\ldots,n\}$. P($i|e^k$) is the probability of slice *i* activated in all the frames of k^{th} category action training dataset.

We define *P*(*i*|*a*), where $i \in \{1,\ldots,m\}$, as ratio of the time the i^{th} slice activated by current action *a* to the time all the slices activated totally, and it can be figured as the contribution of slice *i* to aid the classification of *a*.

Thus, we can calculate the probability of *a* being classified to k^{th} action category predefined. Then, the action category *C* is formulated by Equation 4, satisfying $C \in$ {1, .., n}·

$$C = \arg\max_{k} \sum_{i=1}^{m} P(i \mid a) P(e^{k} \mid i) \tag{4}$$

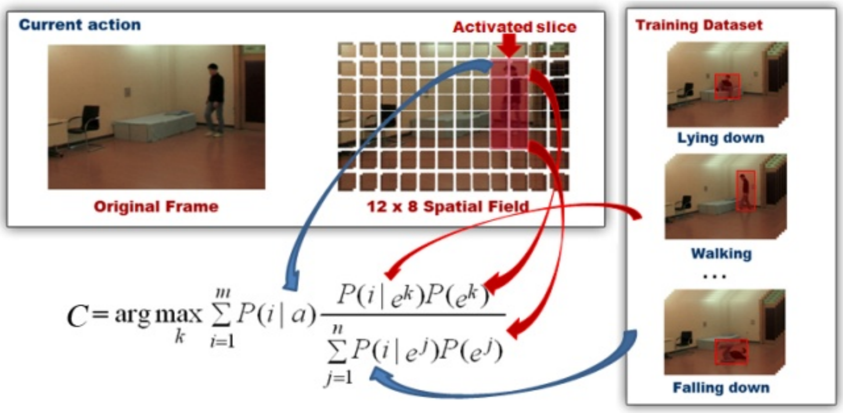

Current action
Activated slice
Original Frame
12 x 8 Spatial Field
Training Dataset
Lying down
Walking
...
Falling down
$C = \arg\max_{k} \sum_{i=1}^{m} P(i \mid a) \frac{P(i \mid e^{k}) P(e^{k})}{\sum_{j=1}^{n} P(i \mid e^{j}) P(e^{j})}$

Fig. 4. The fusion mechanism of our system

Semi-Supervised Learning: We present a semi-supervised mechanism in the classifier labeling step of our system. The flowchart of the mechanism is shown in Figure 5.

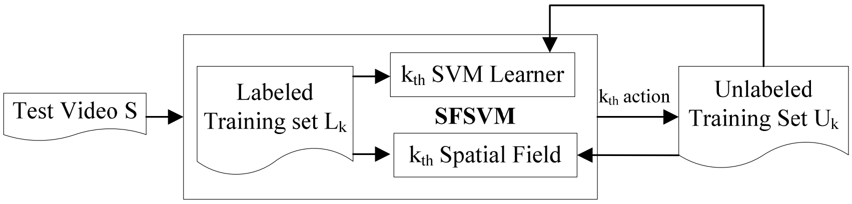


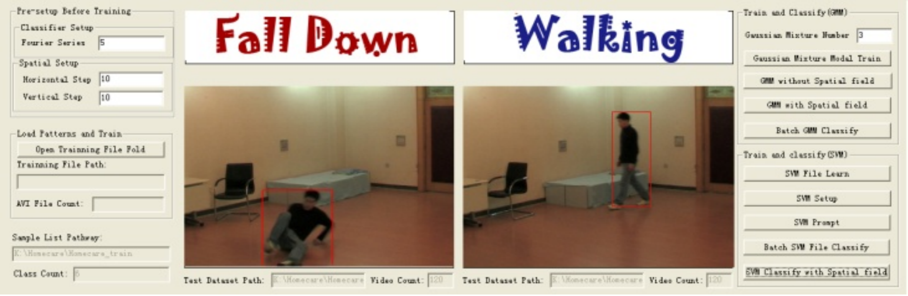
Fig. 5. Semi-supervised incremental training mechanism

In our mechanism, the SFSVM classifier will be updated by both of the labeled and unlabeled samples. Each test sample will be included to unlabeled set, and then classified by SFSVM as k^{th} action. The new training set updates the k^{th} SVM classifier and Spatial Field. Then, to prevent the over fitting problem, we present a 5-fold cross verification strategy: we select a fold by turn from the initial labeled training dataset as validation dataset, and evaluate the effect of classifier trained by both of labeled and unlabeled samples. If the precision on validation set declined, the training procedure stops.

Our mechanism advances on two aspects: (1) The SVM classifier is updated incrementally by the unlabeled samples, which not being classified by SVM but SFSVM for the precision of latter is much higher. This makes the SVM classifier to require less training samples to acquire ideal training effect, which is large cost for human labeling. (2) Include testing samples to unlabeled dataset because the testing video may keep increasing in practical application but the original unlabeled samples will not in our system. The experimental result is very satisfying for the failure rate of this advanced semi-supervised mechanism is 16 out of 70.

4 Experiment

Because there are no public dataset aiming at homecare surveillance, we present a novel test and training data for the experiments of this paper and hope to contribute to other related researchers. The dataset is derived from the video recorded from home environment for twelve hours totally by wide-angle camera. You can download the database from the URL: ftp://vilab.hit.edu.cn/homecare_dataset or send email to the author. Within the video, an actor was instructed to perform a series of activities in the room designed to emulate aspects of the way an older person might use such a room. We define 5 kinds of normal actions and 2 exceptional ones in the experiment which have been defined in section 1. The "Fall" class contained sequences in which the actor was instructed to simulate a fall. (There are obvious barriers to obtaining a video data set of older people falling in reality), and the "Twitch" class was acquired by actor quivering in high speed. 280 sequences for training of 7 action categories are extracted from the whole video including 40 for each respectively which remains 20 unlabeled, and the scale of test dataset involves 30 actions for each predefined category.



Fourier Series: 5

Horizontal Step: 10

Vertical Step: 10

Trainning File Path:

AVI File Count:

Sample List Pathway: K:\Homecare\Homecare_train

Class Count: 6

Test Dataset Path: K:\Homecare\Homecare Video Count: 120

Test Dataset Path: K:\Homecare\Homecare Video Count: 120

Gaussian Mixture Number: 3

Fig. 6. System User Interface

Two experiments are implemented: the first is to compare the effect of action classification on semi-supervised SVM and semi-supervised SFSVM. To scale the performance qualitatively, the values of Precision (*P*) are calculated for evaluation. The second experiment is to compare the performance of our system with the system proposed by Lin [1].

4.1 Action Classification

This experiment is performed for classifying 210 sequences into 7 action categories by pre-trained SVM model whose negative samples contain 7 randomly selected from the sequences of all the other action categories respectively. For the uncertainty of training negative samples, we acquire the result by conducting the experiment for ten times including training and classifying, and average the result of all. Table 1 summarizes the experimental results over both SVM and SFSVM.

The reason of incorrect recognition of SVM is that the surveilled may conduct a specific action in various directions, and some action categories exhibit high similarity in *T-R* curve between each other which will make the Fourier fitting strategy misreport the motion characteristics especially between the action “Lying down” and “Falling down”.

Table 1. Comparison between SVM and SFSVM

Precision of SVM/SFSVM	a1	a2	a3	a4	a5	a6	a7
a1	**.55/.76**	.00/.00	.00/.00	.00/.00	.45/.24	.00/.00	.00/.00
a2	.00/.00	**1.0/1.0**	.00/.00	.00/.00	.00/.00	.00/.00	.00/.00
a3	.00/.00	.00/.00	**.49/.94**	.00/.00	.00/.00	.42/.06	.09/.00
a4	.00/.00	.05/.00	.00/.00	**.95/1.0**	.00/.00	.00/.00	.00/.00
a5	.38/.12	.00/.00	.00/.00	.00/.00	**.62/.88**	.00/.00	.00/.00
a6	.00/.00	.00/.00	.28/.00	.00/.00	.00/.00	**.72/1.0**	.00/.00
a7	.15/.00	.00/.00	.11/.21	.00/.00	.12/.00	.00/.00	**.62/.79**

Note: a1. Squatting down, a2. Standing up, a3. Lying down, a4. Walking, a5. Sitting down, a6. Falling down, a7. Twitching.

Comparing with the results based on SVM, the result of SFSVM is more satisfactory because the classification is optimized by combining Spatial Field with Bayesian. A small quantity of incorrect classification is caused by the Spatial Field over fitting for some actions from different categories may occur in same location especially between lying down and twitching whose Spatial Field are covering each other. Additionally, still some false annotated samples produced by incremental learning mechanism misadjust the interface of different action categories and renews Spatial Field for specific categories falsely.

4.2 Comparison with Former Work

This subsection compares results of our location extraction method with Lin [1], a vertical histogram based fall down detection method. The summary of the evaluation results in Figure 7 shows that our strategy outperforms Lin when the suitable spatial divided strategy is selected. The selection of foreground from Lin’s strategy was done using a confidence threshold to filter the noise. Our improvement could be concluded from: (1) More robust tracking algorithm; (2) Classification mechanism depending less on rules; (3) Integrate motion environment to aid classification.

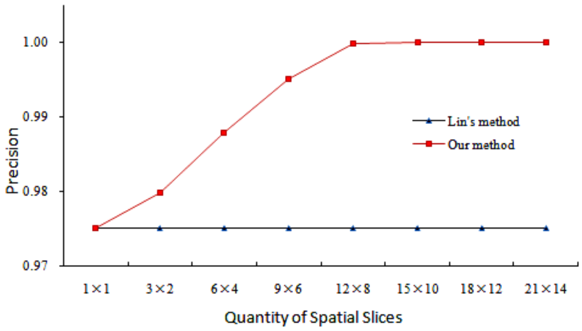


Fig. 7. Experimental measurements of fall down detection.

To analyze the performance of each component in our system, we substitute one component of DCRF-Fourier-SFSVM by others each time. The experiment result demonstrates our framework is the best choice at present, shown at Figure 8.

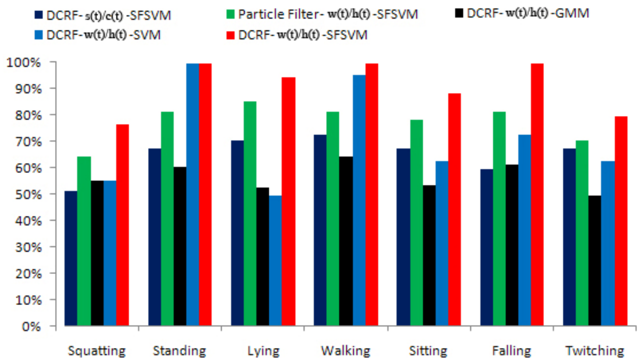


Fig. 8. Framework demonstration result

Note: $w(t)$ and $h(t)$ represent tracking window's width and height in moment t respectively, and $s(t)$ and $c(t)$ represent tracking window's area and perimeter in moment t.

5 Conclusions

A new vision-based semi-supervised homecare system has been proposed in this paper. We improve state-of-the-art vision-based homecare with following contributions: Semi-supervised SVM is employed to recognize multi-actions instead of the rule-based strategy in former papers. We improve the effectiveness of multi-category action classification by introducing the Spatial Field motion constraints with Bayesian inference. We also propose a new motion descriptors based on Fourier fitting to improve the traditional direction-histogram-based descriptor in former action classifier. The experimental results are much more promising comparing with rule-based strategies.

The pair-wise relationship between temporal information and specific action category will be paid more attention. Additionally, since SFSVM is sensitive to the change of environment, and should be retrained in new situation, we will concentrate in the model adaption issue in our future work.

Acknowledgements. This work is supported by NSF of China (No.60775024), State 863 High Technology R&D Project of China (No. 2006AA01Z197), and Program for China New Century Excellent Talents in University (NCET-05-03 34).

References

1. C.-W. Lin, Z.-H. Ling: Automatic Fall Incident Detection in Compressed Video for Intelligent Homecare. ICCCN 2007. Proceedings of 16th International Conference on 13-16 Aug. pp. 1172-1177.
2. C.-W. Lin, Z.-H. Ling, Y.-C. Chang, C. J. Kuo: Compressed-Domain Fall Incident Detection for Intelligent Home Surveillance. ISCAS 2005. IEEE International Symposium on 23-26 May 2005. pp. 3781 - 3784 Vol. 4.
3. N.-C. Hammadi and S. J. McKenna: Activity Summarisation and Fall Detection in a Supportive Home Environment. IEEE Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04). pp. 323-326 Vol.4.
4. S. Bonner: Assisted Interactive Dwelling House: Edinvar Housing Association Smart Technology Demonstrator and Evaluation Site. In Improving the Quality of Life for the European Citizen (TIDE), pp. 396–400.
5. S.-Y. Yang, and C.-T. Hsu: A Study of Moving Objects Extraction for Surveillance Videos by Background - Subtraction Method, Master Thesis. National Tsing Hua University, 2006.
6. V. Vapnik: The nature of statistical learning theory, Springer-Verlag, New York, 1995.
7. D. Chen, J. Yang, R. Malkin, and H. D. Wactlar: Detecting Social Interactions of the Elderly in a Nursing Home Environment, ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 3, No. 1, Article 6, Publication date: February 2007.
8. D. Berrada, M. Romero, P. G. Abowd, Marion Blount, PJohn Davis: Automatic Administration of the Get up and Go test, Jun. 2007. Proceedings of the 1st ACM SIGMOBILE international workshop on Systems and networking support for healthcare and assisted living environments. pp. 73-75.
9. A. M. Tabar, A. Keshavarz, H. Aghajan: Smart Home Care Network using Sensor Fusion and Distributed Vision-Based Reasoning, 2006. Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks. pp. 145-154.